



Better

Evidence-based Education

ASSESSMENT

THE UNIVERSITY *of York*
Institute for Effective Education

JOHNS HOPKINS
UNIVERSITY

School of Education

Center for Research and Reform
in Education

Providing feedback that works

**Promoting learning and achievement
through self-assessment**

Using multiple measures in the classroom

Changing education through research and discovery



We believe that all children and adults deserve proven programs and practices in teaching and learning. In keeping with the Johns Hopkins tradition of research and excellence, our school is committed to building an evidence base for practice that enriches peoples' lives by:

- Researching and disseminating what works in teaching and learning
- Building the next generation of action-oriented researchers
- Providing courses for educators to empower them with evidence-based practices
- Transforming education policy and practice

Join us!

As a Researcher

We are seeking world-class scholars who combine advanced methodological skills with a concern for the practice of education.

As a Student

We offer Ph.D. and Master's courses for individuals who want to transform education through research, teaching, and leadership.

As a Supporter

If you support the goal of evidence-based reform in education, we seek your involvement.

To learn more about the Johns Hopkins School of Education contact us at <http://education.jhu.edu/>.

Editorial

AS CONGRESS ONCE AGAIN takes up the reauthorization of the Elementary and Secondary Education Act (ESEA), the largest federal education program for K–12 schooling, it should think about how to make certain that proven educational practices become widely and effectively used. The best way to produce substantial improvements in student learning on a large scale is through development and rigorous evaluation of promising methods and then scaling up the ones that work.

A model for evidence-based reform in education is the Obama Administration's Investing in Innovation (i3) initiative. It provides large grants to help proven programs that have been repeatedly and rigorously evaluated and found to be effective, to scale up their impacts. The i3 model should be at the core of the reauthorization of ESEA. In particular, Title I, the \$14 billion program to help high-poverty schools, needs to be focused on helping school leaders learn about and then adopt proven approaches to increase achievement in high-poverty schools. We know a lot about what works in education, as evidenced by the articles right here in *Better* magazine. Past issues have highlighted proven approaches for improving achievement in subjects such as reading and math, and the current issue focuses on effective assessment methods. The reauthorization of ESEA gives us an opportunity to put this know-how to work in our schools.



Robert Slavin
Editor-in-Chief
Director of the Center for Research and Reform in Education
Johns Hopkins University

Better: Evidence-based Education is published three times a year by the Center for Research and Reform in Education at Johns Hopkins University and the Institute for Effective Education, University of York, Heslington, York YO10 5DD ©IEE 2011

Phone: 410-616-2444 Email: thebee@bestevidence.org

UK Editor: Jonathan Haslam UK Writer: Jeannette Bollen-McCarthy

US Editor/Writer: Beth Buckheit

Design: Cambridge Publishers Limited ISSN: 2041-921X

The views and opinions expressed in *Better* are those of the authors, and do not necessarily reflect those of the University of York, Johns Hopkins University, or our sponsors.

Sign up for future copies of *Better* at www.bestevidence.org/better



Carol Ann Tomlinson
William Clay Parrish, Jr. Professor and Chair of Educational Leadership, Foundations, and Policy at the University of Virginia's Curry School of Education.



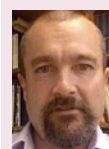
W. James Popham
Professor Emeritus at UCLA Graduate School of Education and Information Studies.



Douglas Fisher
Professor of Teacher Education at San Diego State University and teacher leader at Health Sciences High and Middle College in San Diego, CA.



Jay McTighe
Educational consultant with an extensive background in professional development, and co-author of 11 books.



Steve Higgins
Professor of Education at Durham University in the UK.



Margaret Heritage
Assistant Director for Professional Development at the National Center for Research on Evaluation, Standards and Student Testing (CRESST) at the University of California, Los Angeles.



Susan M. Brookhart
Educational consultant, and author or co-author of 11 books and over 50 articles.



Albert Oosterhof
Professor Emeritus in Educational Psychology and Learning Systems at Florida State University.



Heidi Andrade
Associate Professor of Educational Psychology and the Associate Dean for Academic Affairs at the School of Education at the University at Albany.



Sandra McNally
Senior Research Fellow at the Centre for Economic Performance at the London School of Economics, and director of the Education and Skills program.

CONTENTS

VOLUME 3, ISSUE 3

- 4–5 Assessment in a differentiated classroom
Carol Tomlinson and Tonya Moon
- 6–7 Formative assessments in high schools
Douglas Fisher and Nancy Frey
- 8–9 Formative assessment and feedback to learners
Steve Higgins
- 10–11 Multiple measures in classroom assessment
Susan Brookhart
- 12–13 Promoting learning and achievement through self-assessment
Heidi Andrade
- 14–15 Exposing the imbalance in “balanced assessment”
W. James Popham
- 16–17 Do we need an assessment overhaul?
Jay McTighe and Grant Wiggins
- 18–19 Formative assessment: An enabler of learning
Margaret Heritage
- 20–21 Upgrading high-stakes assessments
Albert Oosterhof
- 22–23 England v Wales: Education performance and accountability
Sandra McNally
- 24–25 Evidence in the news
- 26–27 Latest research

We are part of a movement.

Promethean is pleased to sponsor The Institute for Effective Education's "Better" to promote effective practices for teachers and learners.

Join over 400,000 members at www.PrometheanPlanet.com



Assessment in a differentiated classroom

Assessment at every stage of instruction can help teachers match teaching/learning plans to students' learning needs.

Carol Tomlinson and Tonya Moon explain

ARCHITECTS ARE ADVISORS in the building process, making a complex system manageable by formulating critical decisions, setting quantifiable results, and working closely with others while encouraging colleagues to employ the industry's best practices. The role of the teacher is not dissimilar. The teacher, too, works in a complex system serving as a change agent for students. An effective teacher makes the curriculum (the "what" of teaching) accessible through appropriate instructional practices (the "how" of teaching).

The teacher's role is made more complex by the reality of student diversity in readiness to learn, language, economic background, culture, motivation, interests, approach to learning, and so on. How well the teacher serves as a change agent for the full range of students in the complex system we call a classroom is both measured and informed by persistent use of assessment. In a differentiated classroom, generation and use of data to inform instruction, as well as to measure the effectiveness of instruction, is a core part of the instructional cycle and is critical to the success of both teacher and students.

The collection and use of assessment data to support differentiation occurs in three stages:

- 1 Planning for instruction, including pre-assessment;
- 2 Implementing instruction, including formative assessment; and
- 3 Evaluating instruction, including summative assessment.

Stage 1: Planning for instruction including pre-assessment KUDs

Fundamental to the success of the teaching/learning process is a teacher's clarity about what students must **know**, **understand**, and

be able to **do** (KUDs) as the result of each segment of learning. KUDs provide teacher and students with clarity about learning targets, and also facilitate alignment of teaching, learning, assessment, and differentiation. KUDs should provide a framework that engages learners and promotes their understanding of key content. Engagement is essential for sustained student motivation.

☞ **Not only is it important that students are clear about criteria that will indicate success on assessments, but also that teachers involve them in developing those criteria.** ☞

Understanding is critical for student retention, application, and transfer of what they learn. KUDs also provide parameters for differentiation, because the goal is rarely different content for different learners, but rather different approaches to, and support systems for, mastering required content.

Pre-assessments

With KUDs clearly established, teachers can create pre-assessments that measure a student's current status with target KUDs and critical prerequisites that teachers might otherwise assume students bring to class from past school experiences. Pre-assessment can also expose students' misconceptions about content, cueing teachers to address those barriers. In addition, pre-assessments can be valuable in revealing students' interests, enabling teachers to make content more relevant, and helping teachers grow their understanding of the range of ways in which their students might approach learning most efficiently.

Pre-assessments can take many forms including journal entries, Frayer diagrams, concept maps, short answer "tests," and interest surveys to name a few. Students should understand that pre-assessments are not graded, but serve the purpose of helping the teacher plan how best to move them forward in the unit of study. Teachers in effectively differentiated classrooms: use pre-assessment data to select materials to appropriately challenge their students; assign students to groups based on readiness, interest, and learning preferences; plan for small group instruction, and so on. Pre-assessments help teachers understand the variety of needs in their classrooms as the study of content begins in order to quickly optimize the match between learner need and teacher instruction.

Stage 2: Implementing instruction and formative assessment Teacher role

Once initial instructional plans are informed through the use of pre-assessment information and a unit of study unfolds, the second phase of data-informed instructional planning begins. The teacher in a differentiated classroom regularly uses both formal and informal formative assessment to chart the progress of each learner, and of the class as a whole, toward achieving the designated goals. As with pre-assessment, alignment of formative or on-going assessment, and the instruction that follows with KUDs, is essential. While formative assessment should rarely be used for grading purposes, it is important for teachers to provide students with specific feedback from on-going assessment, in order both to help them understand their own progress and to more readily and accurately contribute to their own academic growth.

On-going assessments can be informal (e.g., observations of or discussions with students as they work, teacher note-taking during small group instruction, asking students to indicate their comfort level with a skill through use of hand signals) or formal (e.g., Frayer diagrams, exit cards, entry cards, writing prompts, systematic use of checklists to monitor student development of knowledge or skill). Data from formative assessments provides a compass for the teacher for forward-planning, in terms of how various students might best access ideas and information, what types of class and homework tasks will serve particular students most effectively at a given time, how to use flexible grouping most effectively, which students need additional support, how to pace instruction, and so on.

ASSESSMENT

The differentiated classroom

Student role

Best practice suggests that the role of students in formative assessment should extend beyond being the subject of observation. Not only is it important that students are clear about criteria that will indicate success on assessments, but also that teachers involve them in developing those criteria so that they are more attuned to and invested in their own success. Likewise, having students evaluate their own formative work according to specified KUDs and carefully developed rubrics can further support learning efficacy. When students get feedback that supports and guides, rather than judges, they are more likely to develop realistic perceptions about their status, and to develop the belief that persistent effort on their part contributes to their success as learners. In other words, formative assessment can and should contribute to what Carol Dweck calls a “growth mindset”—the belief that people can make themselves smarter and more successful through sustained effort.

Stage 3: Evaluating instruction and summative assessment

There are times when teachers seek evidence of understanding in a summative way. Summative assessments are graded, and appropriately occur at transitional points, such as the end of a unit, the end of a marking period, or the end of a semester. Once again, summative assessment should be tightly aligned to the specified KUDs that have guided curriculum design, pre-assessment, instructional decisions, and formative assessment. Summative assessment data helps students benchmark their growth and allows teachers to determine mastery of content. While summative assessment has a “final” aspect, it too is somewhat formative in nature. Teachers can look for patterns in achievement that suggest the need for modification the next time they teach the content. Students can look ahead from summative assessment to the next opportunities to apply skills or understandings they have—or have not—mastered.

Differentiation

Differentiation can play a role in summative assessments, whether they are closed or performance based. The purpose of an assessment is to reveal what a student knows, understands, and can do related to a set of clearly defined objectives (KUDs). If an English language learner understands the process of photosynthesis but is unable to demonstrate that mastery in a test because it requires an essay response, the test has



What we know

- Pre-assessment and formative assessment are critical tools to design instruction that addresses varied learner needs.
- Student achievement benefits when teachers use pre-assessment and formative assessment data to plan instruction to address learner needs.
- Student investment in learning and achievement benefit when students know the learning goals and use feedback from formative assessment to achieve those goals.

failed to reveal what the student knows. It would have been helpful to offer the option of writing an essay, completing a structured chart, or presenting a series of annotated storyboards.

Summative assessments can be differentiated in terms of complexity of the language of directions, providing varied options for expressing learning, degree of structure vs. independence required for the task, the nature of resource materials, and so on. What cannot be differentiated is the set of criteria that determine success. In other words, with the exception of students whose educational plans indicate otherwise, the KUDs established at the outset of a unit remain constant for all students.

Conclusion

The teacher in a differentiated classroom is an architect of instruction designed to maximize the success of all learners. Assessment allows the teacher to be a successful change agent,

working on behalf of students to understand their development as learners and to use best practices to scaffold their growth from varied points of readiness, interest, and approach to learning.

About the authors

Carol Ann Tomlinson is William Clay Parrish, Jr. Professor and Chair of Educational Leadership, Foundations, and Policy at the University of Virginia's Curry School of Education. She was also a public school educator for 20 years. Her work focuses on differentiated instruction.

Tonya R. Moon is a Professor at the University of Virginia's Curry School of Education. Her interests and work focus on student assessment and the development of talent in school age children.

Further reading

Graham S (2010), *Teaching Writing*.

Black P and Wiliam D (1998), *Inside the Black Box: Raising Standards Through Classroom Assessment*. Phi Delta Kappan, 80, 139–148.

Dweck C (2006), *Mindset: The New Psychology of Success*. New York: Ballentine.

Earl L (2003), *Assessment as Learning: Using Classroom Assessment to Maximize Student Learning*. Thousand Oaks, CA: Corwin.

Stiggins RJ, Arter JA, Chappuis J, and Chappuis S (2004), *Classroom Assessment for Student Learning. Doing it Right – Using it Well*. Portland, OR: Assessment Training Institute.

Formative assessments in high schools



Douglas Fisher and Nancy Frey describe a four-step approach to successful formative assessment

TEACHERS WHO USE FORMATIVE

ASSESSMENTS target their instruction to student needs. Teachers who do not use formative assessments rely on guesswork or generic plans, rather than basing their instructional decisions on data. The difference in student learning is striking. Simply said, the systematic use of formative assessments results in better learning.

As a quick comparison, formative and summative assessments differ in terms of when they are administered, how often they are used, and what is done with the results (see Figure 1). While both are useful in teaching and learning, our focus here is

on formative assessments that are useful in planning instructional interventions.

Formative assessments work when they are systematic and purposeful. The unfocused use of random data collection tools cannot yield focused instruction. Teachers have a lot of data about students, some of which is useful in planning and some of which is not. What teachers need is a system for effectively using student performance data to make instructional decisions, rather than another test. We have developed such a system with four components: feed-up, checking for understanding, feedback, and feed-forward.

Figure 1

Differences	Formative assessments	Summative assessments
When administered	Daily, if not more often	At the conclusion of a unit of study
Frequency of use	Ongoing, regularly	Typically once or twice per unit of study
Use of results	Inform teacher about effectiveness of lessons; plan instruction; identify students in need of immediate additional support	Grades; accountability systems; inform teacher about effectiveness of instruction or intervention; identify students in need of long term support

Feed-up

As part of any formative assessment system, both teacher and students have to understand the purpose of the lesson and the purpose of the assessment. A clear purpose should be communicated with students on a daily basis. In addition to staying focused on the topic, a clear purpose allows the teacher to determine when learning targets have been met. For example, consider the following two purpose statements:

- To understand the causes and effects of World War II.
- To explain why the attack on Pearl Harbor resulted in the United States' entry into World War II.

The second highlights what students should learn on a specific day and provides the teacher with information about what to assess. With a clear purpose, communicated with students, teachers are ready for the second component.

Checking for understanding

There are a number of excellent ways that teachers can check for student understanding, including:

- *Oral language*, such as inviting students to explain how they solved an Algebra problem, putting their thumbs up or down to indicate agreement with a prediction during a science experiment, or holding up a yes or no response card when asked about a character's actions in a play.
- *Questions*, including those that require critical thinking such as "compare and contrast the roles of Snowball and Napoleon in *Animal Farm*." Questions can be oral or written and may include a range of student responses from recall to synthesis.
- *Writing*, such as "exit slips" in which students summarize their understanding of sources or respond to prompts such as: "How are we connected to our environment?"
- *Projects and performances*, through which students demonstrate their understanding, such as an erosion map in Earth Science or an enactment of a play.
- *Quizzes and tests*, which are used to plan instruction rather than provide students with a grade. In the case of formative assessments, each of the distracter items should be diagnostic in terms of student

Figure 2: Error analysis in social sciences Date: 10/12 Topic: Reading primary sources

Error	Period 1	Period 2	Period 3	Period 4	Period 5
Skimming and scanning; preview text	JC			AA	
Predicting errors	JC, JT, AG, DL, TV	EC, MV, WK		AA, SK, MG, EM, BA, TS	HH, DP, MR, CH
Making connections	JC, AG, SL	WK, MW		AA, BA	MR
Sourcing	JC, JT, DL, MM, SL, ST, ND	RT, VE, VD, CC		AA, MG, SC, PM, LG	DP, DE
Drawing conclusions	JC, JT, MM	EC, SJ		AA, MG, BA, GL, PT, DO, DE, LR, SK, EM, TS, LG, PM, DP, RT, HA, KJ, DE, RC, DW, DL, KS, IP, SN, MW, JG, KE, JV	DE, MR, DC, AT

understanding. For example, one of the wrong choices for an Algebra problem might be based on choosing the wrong algorithm, whereas another choice might be based on common mathematical errors such as forgetting to change the number from negative to positive when squaring.

Feedback

When the teacher has formative assessment data, students expect feedback. For feedback to be effective, it must be timely, specific, and understandable. However, even when those conditions are met, feedback by itself has limited impact on student understanding and thus achievement. If you doubt this, check the trash can (or recycling bin) following a lesson, and you will find examples of student work on which the teacher has written a great deal of useful information. The teacher has worked hard, providing students with information about their performance, but it has gone to waste.

In a formative assessment system, students receive feedback that is more global in nature. The specifics are left for feed-forward, which we will focus on next. For example, on an essay, a student received the following comment from his teacher: “Your transitions and organization have significantly improved. The next area of attention might be word choice. You express your ideas well and could make a stronger case with increased academic vocabulary.” In addition, the student received a rubric with specific areas identified and a grade on the paper. This feedback focused on what the student had demonstrated successfully, based on previous instruction, and what the student might focus on next. In addition, the rubric provided feedback about specific aspects of the paper, such as mechanics, thesis development, and so on. Of course there was also a summative component and a grade.

Feed-forward

Unfortunately, too many teachers use feedback as their entire formative assessment system. This transfers all of the

What we know

- Formative assessment systems are based on student performance and provide teachers with data about what to teach next.
- Feedback is not effective in isolation; rather it needs to be acted on by both students and teachers.
- Checking for understanding, based on a clear purpose, is an important part of a formative assessment system and should not be overlooked.

responsibility for learning to the student and fails to provide the teacher with specific teaching points. Remember that the power of a formative assessment system is that it can inform the teacher about next steps for instruction. The feed-forward element of this formative assessment system ensures that data is collected and analyzed for patterns so that the teacher can make informed decisions about what should be taught next.

An error analysis sheet is a vital tool for feeding forward. When teachers provide students with targeted feedback, as in the example above, they have time to focus on the errors students make. But the next step is to record this feedback for analysis. For example, the error analysis sheet a social sciences teacher created as she was teaching her students to read primary source documents (see figure 2). The teacher identified errors her students were making, listed them on a code sheet, and entered students’ initials in the corresponding place on the table. In this way, she was able to focus on trend data, which gave her an actionable plan. It is clear that fourth period needs a great deal more instruction on drawing conclusions. This is considered a global error, given that more than 60% of the students in the class made this same error. The other identified errors are targeted errors and the students who made them need additional small group support. The trend data also identifies students, such as JC in first period and AA in fourth period,

who need intensive interventions if they are to be successful in this class. This analysis of student work has led the teacher to specific actions they can take to plan instruction and meet student needs.

Conclusion

The error analysis sheet is a concrete example of the ways in which a formative assessment system can be used to plan instruction. Having said that, it is important to note that this was not done in isolation. The students and their teacher understood the purpose of the lesson and the assessment, the teacher checked for understanding throughout the lesson, and collected data to analyze. Had she simply identified errors, nothing much would have changed in terms of student understanding. The feed-forward aspect of the formative assessment system provided the teacher with information about which students needed to be taught what next. Taken together, the formative assessment links student work with instruction such that better learning occurs.

About the authors

Douglas Fisher and **Nancy Frey** are professors of teacher education at San Diego State University and teacher leaders at Health Sciences High and Middle College in San Diego, CA. They focus their research on quality teaching, including the use of assessment information to inform instruction.

Further reading

Fisher D and Frey N (2007), *Checking for Understanding: Formative Assessments for Your Classroom*. Alexandria, VA: ASCD.

Frey N and Fisher D (2011), *The Formative Assessment Action Plan: Practical Steps to More Successful Teaching and Learning*. Alexandria, VA: ASCD.

Hattie J and Timperley H (2007), The Power of Feedback. *Review of Educational Research*, 77, 81–112.

Formative assessment and feedback to learners

Feedback to students is at the heart of successful teaching, but research suggests that *how* this is given is key to whether it is effective. **Steve Higgins** explains

I WORK CLOSELY WITH TEACHERS and am passionate about supporting them with research evidence that is helpful and practical for their teaching. Recently I have been undertaking a review of the evidence about what works for learners for the Sutton Trust, a UK charity which aims to improve

educational opportunities for children and adolescents from non-privileged backgrounds and to increase social mobility. One of the questions that teachers have frequently asked me for the review, is about what works in terms of formative assessment and feedback to students.

Assessment and learning

Feedback is an essential part of the learning process, but both students and teachers are often disappointed or even frustrated at the feedback process. Students complain they don't know what to do when they get the results of assessments, or even say getting feedback is demoralizing. More critically they often say that feedback comes too late to be of any use to them at all.

One of the aims of assessment and testing of students in the classroom is that it should help teachers teach more effectively, by understanding what their students already

☞ An assessment activity can help learning in this way only if it provides information that teachers and their students can use as feedback in assessing themselves and one another ☞

know or can do. It should also help students understand what they have to do next to improve their own learning.

Formative assessment and 'assessment for learning'

In recent years there has been increasing interest in formative assessment where information is used by the teacher or by the learner as information to change what they do next in a teaching or learning activity. Assessment for Learning is an assessment task in which the main purpose is to promote or improve students' learning. This is different from assessments that aim to hold schools or teachers accountable or to identify the competence or ranking of students. An assessment activity can help learning in this way only if it provides information that teachers and their students can use as feedback in assessing themselves and one another, and then in modifying the teaching and learning activity. Assessment for learning only becomes "formative assessment" when it leads to this change.

We know that frequent summative tests and assessments have a negative

Feedback about:	Examples	Key points
The task	Feedback about how well the task is being achieved or performed, such as: <ul style="list-style-type: none"> ● Indicating where correct responses are different from incorrect. ● Getting more or different information relevant to the task. ● Building more task knowledge. ● Prompts and direct cues. 	Feedback that focuses even more on correct than incorrect behaviors, and which encourages the learner. Being positive about errors as learning opportunities.
The process	Feedback specific to the processes of learning, the <i>how</i> rather than the <i>what</i> , or relating and extending tasks such as identifying: <ul style="list-style-type: none"> ● Connections between ideas. ● Strategies for spotting mistakes. ● Explicitly learning from mistakes. ● Cues to the learner about different strategies and errors. 	Identifying where in the process to focus attention to improve relative to previous attempts.
Self-regulation	How students monitor, manage, and regulate their actions towards the learning goal, such as their: <ul style="list-style-type: none"> ● Capability to identify feedback themselves and to self-assess. ● Willingness to put effort into seeking and dealing with feedback. ● Having confidence they are correct. ● Positive attributions about success AND failure. ● How good they are at help-seeking. 	Needs to emphasize success at challenging activities through effort, focusing on specific strategies for self-regulation which led to their success. Corrected errors are a key part of this.
The individual	Praise directed to the effort, self-regulation, engagement, or processes relating to task/performance: <ul style="list-style-type: none"> ● e.g., "You're really great because you have worked hard to complete this task by applying this concept" NOT "good girl." 	The most common, but most dangerous kind of feedback. Tends to be too general and personal. Feedback should rather emphasize what the individual has done (or could do), not who they are.

ASSESSMENT

The importance of feedback

impact on students' views of themselves as learners. This is especially true with "high-stakes" testing, when teachers may narrow the curriculum they teach to match the test. This suggests that such assessments are more important for school or teacher accountability than for learning.

We know too that simply practicing assessments will improve students' performance, at least in the short term, but this does not help them with their learning. It's rather like squeezing a child's balloon, the bulge you make when you squeeze it makes the balloon look like it is getting bigger, but there is really no more air in there. Once you let go, it goes back to the size it was before. Test practice is a bit like this in that the students aren't learning anything new. You are just squeezing the balloon. The way you get more air in the balloon is through more effective instruction. A key component of this is feedback which keeps teaching and learning on track to achieve its goals.

This suggests a closer examination of feedback is needed. The analysis that follows focuses on what teachers can do in terms of *how* they give feedback to learners and *what* they get students to think about, rather than other parts of the feedback cycle (such as how they might alter their instruction, or how learners can give feedback to each other). It is based on a number of reviews, but in particular John Hattie's analysis. His work indicates that there are different kinds of feedback to consider. These are about the **task** itself, about the **process** of the task or activity, about students' management of their own learning or their **self-regulation**, and about them as **individuals** and who they are. Research suggests that feedback is best directed at the first three levels. In addition, evidence shows that:

- It should be about *challenging* tasks or goals (rather than easy ones);
- It is even more important for teachers to give *feedback about what is right* rather than what is wrong;
- Feedback should be as *specific* as possible and, ideally, compare what students are doing *right* now with what they have done *wrong* before; and
- It should *encourage* students, and not threaten their self-esteem.

Conclusion

Successful feedback is that which leads to action on the part of the teacher or learner and closes the formative assessment loop. Teachers should be very specific about their feedback and what to do in response, and



What we know

- Feedback is central to the teaching and learning process and keeps it on track.
- It closes the loop between Assessment for Learning and formative assessment by enabling action by the teacher and/or learner.
- Letting students know when they get things right, and *why* they are correct is even more important than pointing out mistakes or errors.
- Specific feedback is more useful than general, particularly where this relates to previous work students have done.
- Praise should be specific to what the student has done.
- Feedback should encourage and not demoralize learners.

should encourage students to see mistakes as opportunities to improve.

About the author

Steve Higgins is Professor of Education at Durham University in the UK. He has an interest in classroom interaction and the role of the teacher in supporting learning and promoting meta-cognition and self-regulation, particularly where technology is involved. (s.e.higgins@durham.ac.uk)

Further reading

Black P, Harrison C, Lee C, Marshall B, and William D (2003), *Assessment for Learning: Putting It into Practice*. Buckingham, UK: Open University Press.

Hattie J (2008), *Visible Learning*. London: Routledge.

Research used in this article

Black P, Harrison C, Lee C, Marshall B and William D (2004), Working Inside the Black Box: Assessment for Learning in the Classroom, *Phi Delta Kappan*, 86 (1) 9–21.

Harlen W, and Deakin-Crick R (2002), *A Systematic Review of the Impact of Summative Assessment and Tests on Students' Motivation for Learning*. In: Research Evidence in Education Library. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Hattie J, and Timperley H (2007), The Power of Feedback, *Review of Educational Research* 77(1) 81–112.

Kluger AN, and DeNisi A (1996), The Effects of Feedback Interventions on Performance: A Historical Review, A Meta-analysis, and a Preliminary Feedback Intervention Theory. *Psychological Bulletin*, 119 (2), 254–284.



Multiple measures in classroom assessment

Susan Brookhart explains the benefits of a ‘multiple measures’ approach to assessment

IN EDUCATION, THE TERM “MULTIPLE MEASURES” means basing a conclusion or a decision on more than one source of evidence. It should not be a surprise that multiple measures are important in education, because they certainly are in other areas. Have you ever been late for a doctor’s appointment, rushed in, and found your blood pressure was high? Would you expect your doctor to prescribe medication for high blood pressure based on that one reading? Or have you ever sat on a committee? Would you expect the committee to decide on its strategy based on the opinion of only one member? Of course not.

But in education, we often have to work hard to prevent people from making judgments based on one test score or

other single piece of information. I hope this article will do two things: First, help you understand – and be able to explain to others – why multiple measures are important, and second, help you learn to use multiple measures for both formative and summative classroom assessment.

Why use multiple measures in classroom assessment?

The reason to use multiple measures in the classroom is to obtain better evidence about whatever you are trying to measure. Do you want to know about Ethan’s understanding of Colonial America? Or are you trying to work out Carmen’s comprehension of the main themes in *To Kill a Mockingbird*? Perhaps you are interested in diagnosing

where Jayden needs help in factoring quadratic equations? These classroom assessment occasions are all better met with multiple measures.

One way in which multiple measures help to obtain better information is that they boost reliability, reducing the chances that your assessment is uncharacteristic in some way (like high blood pressure after rushing to the doctor). Having more than one measure helps to even out little anomalies in performance that happen by chance. And reliability contributes to validity. There is obviously no way that evidence about achievement can be sound, can be valid, and can be the basis for well-founded decisions, if it is unreliable.

Another way in which multiple measures help improve validity is by enriching the limited picture that any one sample of test questions or performance tasks gives. Carmen is reading *To Kill a Mockingbird*,

and you need evidence of the sense she is making of this complex piece of literature. What are the themes; how did the author use plot, character, and setting together to create them; does Carmen think any of these themes are important today? A set of test questions, a paper, and participation in a prepared debate together allows for a broader set of evidence about the wide variety of knowledge and skills involved in “understanding the themes in *To Kill a Mockingbird*” than any one of these measures alone.

Multiple measures for formative assessment

Formative assessment is a systematic process through which teachers and students gather evidence of learning with the express goal of improving achievement. The assessments used for this are typically informal classroom tests. These help with both teacher feedback and student self-assessment, and result in specific information to inform forward-planning.

An excellent way to incorporate multiple measures in formative assessment is to use a series of assessments, each based on one or more aspects of the ultimate learning goals. These would not be graded, but generate feedback from teacher, peers, and/or self-evaluation. For certain kinds of tasks, this might be a series of similar assessments. For example, students learning to write descriptive paragraphs

Multiple measures help improve validity by enriching the limited picture that any one sample of test questions or performance tasks gives

might write three of them, with the first two serving as practice. Based on the feedback they received each time, and on their own reflection, the final paragraph might result in a grade (summative assessment).

For other kinds of tasks, this might be a series of assessments giving students and teachers information on different aspects of the learning goals. Again, these would not be graded, but would generate feedback from teacher, peers, and/or self-evaluation. For example, as part of their study of Colonial America, students might write or answer questions about the founding of the English colonies, their various government structures and functions, trade and commerce, and so on. After receiving feedback (including self-reflection) on these various aspects of the unit, students should be equipped to study better for a unit test and do better on graded projects or performance assessments.

What we know

- Used wisely, multiple measures in classroom assessment yield richer, more accurate evidence about student achievement than single measures alone.
- For formative assessment of a learning goal, use a series of practice assignments or assessments over various aspects of the learning domain. Make sure students receive feedback (from teacher, peer, or self) that moves learning forward.
- For summative assessment of students' level of achievement of a learning goal, use several assessments that, taken together, better represent the expectations of the learning domain than one assessment could alone.

Multiple measures for summative assessment

In the Colonial America example, a final set of summative, graded assessments might include a test, project, and performance assessment. Unit goals for understanding and being able to explain the importance of historical figures, events, and developments in Colonial America describe a multi-faceted domain of learning. A unit test would be particularly suited to capturing knowledge of facts and concepts and the kind of thinking and application that brief essay questions can measure. A project, perhaps a term

learn than one assessment would yield. If the teacher averages the grades, using a mean or median, this is a *compensatory* use of multiple measures. That is, high performance on one assessment can compensate for low performance on another assessment.

Sometimes teachers differentiate assessment, allowing scores on one assessment to substitute for scores on another. This is a *complementary* use of multiple measures, and although it can be useful, it can easily be misapplied as well. Suppose the teacher in our Colonial America example knew a certain student had a debilitating fear of speaking in public. For that student, she might well substitute (a complementary use of multiple measures) an essay arguing the point-of-view the student would have been assigned in the debate. This is a wise and fair accommodation, since speaking in public is not the learning goal.

However, suppose the teacher decided she would allow students to pick the best one of their three grades (test, project, debate) and use that for the whole unit grade. If the goal is a grade that accurately reflects the student's level of achievement for the unit, this complementary use of multiple measures is *not* wise or well-founded. Information about different aspects of the learning goals is lost. In effect, the goal here becomes to use the highest grade, without regard to the information it conveys.

About the author

Susan M. Brookhart, PhD, is an educational consultant. Her interests include formative and summative classroom assessment and the connection between classroom assessment and large-scale assessment. She is author or co-author of 11 books and over 50 articles on these topics.

Further reading

Brookhart SM (2009), The Many Meanings of “Multiple Measures.” *Educational Leadership*, 67(3), 6–12.

Brookhart SM (forthcoming), *Grading and Learning: Practices that Support Student Achievement*. Bloomington, IN: Solution Tree.

Bernhardt V (1998, March), *Multiple Measures*. Invited Monograph No. 4, California Association for Supervision and Curriculum Development. <http://eff.csuchico.edu/downloads/MMeasure.pdf>

Williamson GL (2006), *Managing Multiple Measures*. Durham, NC: MetaMetrics. www.fcrr.org/fair/Managing_Multiple_Measures-Sept.pdf

Promoting learning and achievement through self-assessment

Self-assessment is a matter of students getting useful feedback *from* themselves, *for* themselves, says Heidi Andrade

RESEARCH HAS SHOWN THAT FEEDBACK can promote learning and achievement, yet most students typically get little constructive criticism on their works-in-progress. This scarcity of feedback is due, in large part, to the fact that few teachers have the luxury of regularly responding to each student's work. Fortunately, research also shows that students themselves can be useful sources of feedback by thinking about the quality of their own work rather than relying on their teacher as the sole source of evaluative judgments.

Student self-assessment is a process of *formative* assessment during which students reflect on the quality of their work, judge the degree to which it reflects explicitly stated goals or criteria, and revise accordingly. Please note the emphasis on the word *formative*. Self-assessment is done on work in progress in order to inform revision and improvement: it is not a matter of having students determine their own grades. Self-evaluation, in contrast, refers to asking students to grade their own work, perhaps as part of their final grade for an assignment. Given what we know about human nature, as well as findings from research regarding students' tendency to inflate self-evaluations when they will count toward grades, I subscribe to a purely formative type of student self-assessment.

The purpose of self-assessment

The purpose of engaging students in careful self-assessment is to boost learning and achievement, and to promote academic self-regulation, or the tendency to monitor and manage one's own learning. Research suggests that self-regulation and achievement are closely related: students who set goals, make flexible plans to meet them, and monitor their own progress are also more likely to learn more and do better in school than students who do not. Self-assessment is a core element of self-regulation because it

involves awareness of the goals of a task and checking one's progress toward them.

The features of self-assessment

Although even young children are typically able to think about the quality of their own work, they do not always do so, perhaps because one or more necessary conditions are not present. In order for effective self-assessment to occur, students need:

- Awareness of the value of self-assessment;
- Access to clear criteria on which to base the assessment;
- A specific task or performance to assess;
- Models of self-assessment;
- Direct instruction in and assistance with self-assessment;
- Practice;
- Cues regarding when it is appropriate to self-assess; and

☞ **The purpose of engaging students in careful self-assessment is to boost learning and achievement, and to promote academic self-regulation, or the tendency to monitor and manage one's own learning** ☞

- Opportunities to revise and improve the task or performance.
- This list of conditions might seem prohibitive but several of the key conditions listed above, including modeling, cueing, direct instruction, and practice, are commonly employed classroom practices. The second condition – access to clear criteria on which to base self-assessment – can be met by introducing a rubric.

There are a number of ways to engage students in effective self-assessment. In general, the process involves the following three steps:

1. *Articulate expectations.* The expectations for the task or performance are clearly

articulated, either by the teacher, by the students, or both. Because students become better acquainted with the task at hand when they are involved in thinking about what counts, I often co-create all or part of a rubric in class by analyzing examples of strong and weak pieces of student work.

2. *Self-assessment.* Students make a first attempt at their assignment, be it an essay, word problem, lab report, fitness plan, speech, or self-portrait. They monitor their progress on the assignment by comparing their work to the expectations articulated on the rubric. In my research on self-assessment of writing, I ask students to use colored pencils to underline key phrases in the rubric, then underline in their drafts the evidence of having met the standard articulated by the phrase. For example, students underline “clearly states an opinion” in blue on their persuasive essay rubric, then underline their opinions in blue in their persuasive essay drafts. To assess one aspect of sentence fluency, they underline “sentences begin in different ways” in yellow on their rubric, use the

same yellow pencil to circle the first word in every sentence in their essays, and then say the circled words out loud with an ear for repetition. If students find they have not met a particular standard, they write themselves a reminder to make improvements in their final drafts. This process is followed for each criterion on the rubric, with pencils of various colors. And it is quick—the whole process can take as little as one lesson, or can be done as homework.

3. *Revision.* Students use the feedback from their self-assessments to guide revision. This last step is crucial. Students are savvy, and will not self-assess thoughtfully unless they know that their efforts can lead to

ASSESSMENT

The benefits of self-assessment



opportunities to actually make improvements and possibly increase their grades.

The value of self-assessment

Some research suggests that simply handing out and explaining a rubric may increase students' knowledge of the criteria for an assignment and help students produce work of higher quality – or it may not. Simply handing out a rubric does not guarantee much of anything. Actively involving students in using a rubric to self-assess their work, however, has been associated with noticeable improvements in students' performances in writing, social studies, mathematics, science, and external examinations. In each case, students were either engaged in written forms of self-assessment using journals, checklists, and questionnaires, or oral forms of self-assessment, such as interviews and student–teacher conferences.

To date, the bulk of the research on self-assessment has been done on writing

and mathematics. Studies of writing have revealed an association between rubric-referenced self-assessment and the quality of writing done by students in elementary and middle school. Of note is the fact that improvements in the conventions of language (grammar and spelling) tend to be negligible: the advances in students' writing were the result of stronger performance on substantive criteria such as ideas and content, organization, voice and tone, and plot development. Similarly, studies conducted in math classes found that students who were taught to self-assess outperformed other students on word problems. In both writing and math, the differences between the treatment and comparison groups were practically, as well as statistically, significant.

Conclusions and encouragement

Student self-assessment can have powerful effects on achievement. The effect can be both short-term, as when self-assessment positively influences student performance on a particular assignment, as well as long-term, as students become more self-regulated in their learning. I encourage educators to take advantage of what we now know about self-assessment by doing the following:

- Articulate the criteria by which students assess their work;
- Teach students how to apply the criteria;
- Give feedback on their self-assessments;
- Give help in using self-assessment data to improve performance;
- Provide sufficient time for revision after self-assessment; and

- Do not turn self-assessment into self-evaluation by counting it toward a grade. Under these conditions, self-assessment can ensure that all students get the kind of feedback they need, when they need it, in order to learn and achieve.

About the author

Heidi Andrade is an Associate Professor of Educational Psychology and the Associate Dean for Academic Affairs at the School of Education at the University at Albany. Her most recent book, co-edited with Greg Cizek, is the *Handbook of Formative Assessment*. (handrade@uamail.albany.edu)

Further reading

Original article, from which this piece was taken: Andrade H, and Valcheva A (2009), Promoting Learning and Achievement through Self-assessment, *Theory Into Practice*, 48(1), 12–19.

Andrade H (Ed.) (2009), Classroom Assessment, *Theory Into Practice*, 48(1).

Andrade H, Du Y and Wang X (2008), Putting Rubrics to the Test: The Effect of a Model, Criteria Generation, and Rubric-referenced Self-assessment on Elementary School Students' Writing. *Educational Measurement: Issues and Practices*, 27(2).

Gregory K, Cameron C and Davies A (2000), *Self-assessment and Goal-setting*. Merville, British Columbia, Canada: Connection Publishing.

What we know

- Self-assessment involves students in thinking about the quality of their own work, rather than relying on their teachers as the sole source of feedback.
- Studies have shown that teaching students to reflect on the quality of their work, judge the degree to which it reflects explicitly stated criteria, and revise, is associated with meaningful improvements in their grades.

Exposing the imbalance in ‘balanced assessment’

There are three elements to ‘balanced assessment,’ but W. James Popham argues that only two deserve their place

IT IS DIFFICULT TO ATTEND any sort of assessment-relevant educational conference these days without hearing someone extol the virtues of “balanced assessment.” In the U.S., what’s typically being described by the proponents of balanced assessment is the application of three distinctive measurement strategies: *classroom assessments*; *interim assessments*; and *large-scale assessments*.

Balanced assessment, as is the case with “balanced” anything, sounds so delightfully defensible. Those who oppose balanced assessment are apt to be the sorts of villains who want “low standards” instead of “high standards” and who applaud “unreliable tests” instead of tests reeking of reliability. Whatever is balanced seems, *a priori*, to be wonderful, but in this case the term may be misleading.

The ‘Blessed Trinity’ of balanced assessment

Briefly, the three measurement strategies of balanced assessment are:

Classroom assessments, typically teacher-made, are currently employed by

most teachers for the purpose of grading their students or as motivators when urging students to “study hard for the upcoming test.” Classroom assessments can also supply timely evidence whenever teachers use formative assessment.

Interim assessments are usually purchased from commercial vendors, but are sometimes created locally. These are standardized tests, typically administered by a district or a state, perhaps two or three times during the school year. Interim tests are intended to fulfill one of the following measurement missions: (1) a predictive function, such as identifying students who are at risk of failing a subsequent high-stakes test, (2) an evaluative function, such as appraising the effectiveness of a recently concluded educational program, or (3) an instructional function, such as supplying teachers with instructionally useful diagnostic data regarding their students. Occasionally, interim tests are intended to supply evidence for more than one of these functions.

Large-scale assessments are almost always created by assessment organizations, either

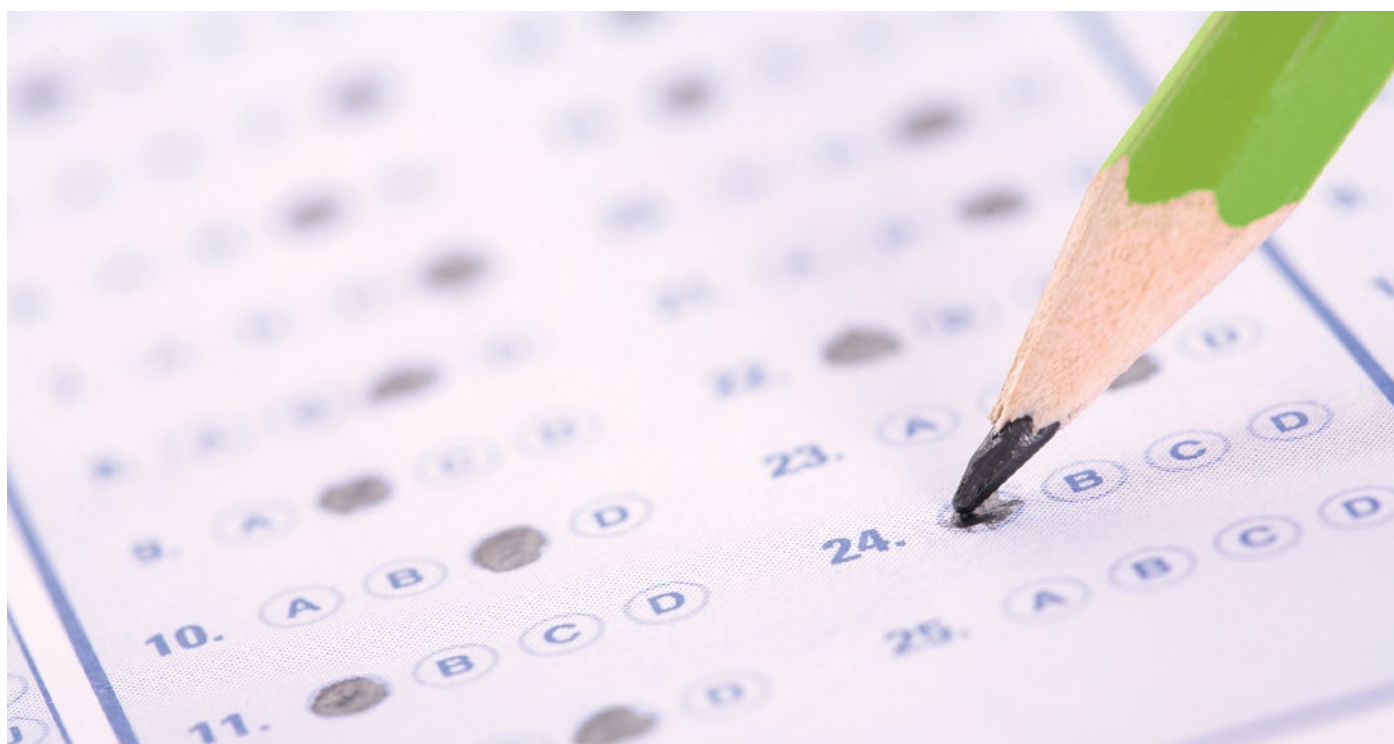
for-profit or not-for-profit groups. In the U.S., the most common examples of these sorts of tests are the annual accountability assessments administered by all U.S. states. Although large-scale assessments are used for purposes other than accountability, for instance, as college entrance exams, the large-scale tests associated with the balanced assessment are typically achievement tests intended for use in an accountability context.

A party crasher

Two of these types of assessment are supported by strong evidence, but one is trying to crash the measurement party without the proper admission credentials.

Classroom assessments

It’s not classroom assessments. Classroom assessments are supported by a formidable array of empirical evidence showing that, when used properly, they trigger substantial growth by students. When classroom assessments are used as part of formative assessment – a process wherein assessment-elicited evidence is used by teachers and/or students to make necessary



ASSESSMENT

Exposing the imbalance

adjustments in what they are doing – there is an abundance of empirical evidence to show that the formative-assessment process is remarkably effective. In their seminal 1998 review of classroom-assessment studies, Paul Black and Dylan Wiliam concluded that formative assessment works conclusively, it produces powerful gains in students' achievement, and it is sufficiently robust so that teachers can use it in a variety of ways, yet still get glittering results. Subsequent empirical investigations continue to support the instructional payoffs of appropriately employed classroom assessments.

Large-scale assessments

Nor are large-scale assessments the party crashers. Large-scale tests, particularly those employed for accountability purposes, enjoy enormous support among both educational policy makers and the public at large. The public are increasingly demanding hard evidence that their schools are being successful, and that their taxes are being well-spent. Not placated by educators' reassurances, educational policy makers at all levels, local to national, are demanding hard, test-based evidence regarding students' achievement. Large-scale accountability tests supply such evidence, and will remain in place until an incredulous citizenry becomes convinced that our schools are working.

⌘ Those who oppose balanced assessment are apt to be the sorts of villains who want “low standards” instead of “high standards” and who applaud “unreliable tests” instead of tests reeking of reliability ☺

Interim assessments

However, in contrast to the other two types of assessment, interim assessments are neither supported by research evidence, nor are they regarded by the public or policy makers as being of particular merit. Indeed, most members of the public and most policy makers don't even know that interim assessments exist.

The chief advocacy for including interim assessments as one of the three strategies of balanced assessment, not surprisingly, comes from the vendors who sell them. Many district-level administrators are desperate to prevent their schools from getting low scores on annual state accountability tests, and so are swayed by the glowing words about the positive instructional payoffs that

What we know

- Classroom formative assessment works well, and the process can be successfully used by classroom teachers in diverse ways.
- Society now demands evidence from large-scale accountability tests to evaluate the success of tax-supported schooling.
- Interim assessments, at the moment, are supported neither by research evidence nor by a societal demand.

accompany commercially peddled interim tests. It is not surprising that many district officials purchase interim assessments for their teachers.

Yet, at the 2010 annual meeting of the National Council on Measurement in Education in Denver, Judith Arter – based on her careful review of research studies regarding interim tests – concluded that no meaningful empirical support currently exists for interim assessments. Regretfully, she noted that “the amount of attention being put on having interim assessments in place saps resources from other formative practices supported by a much larger research base.”

Accordingly, when it comes to the support associated with these three assessment approaches, one of them is blatantly out of balance with its assessment cousins.

A serious shortcoming?

Interim tests, other than being seen by some armchair analysts as “rounding out” the balanced-assessment picture, come to us without compelling support, either empirical or political. In the U.S. where almost any TV-advertised health product is accompanied these days by an allusion to “clinical evidence” supporting the product's virtues, the promotional literature accompanying America's interim assessments is particularly light on evidence, of any sort, that they are worth what they cost. And their costs are not trivial, either in terms of money spent or in classroom time taken.

Perhaps, in the future, research evidence supporting the instructional dividends of interim assessments will be available. However, it's possible that there is an inherent, but unrecognized flaw in the interim-assessment approach, a flaw that dooms these tests to be ineffectual, particularly in improving instruction. Interim tests, you see, are administered at a given time during the school year, for instance, in the middle of or near the close of every three-month segment. So, in order for the results of these tests to help teachers instructionally, the timing of the teacher's instruction must mesh with

what's covered in a given interim test. A test covering yet-untaught content, or content that was treated weeks ago, will hardly inform a teacher's decision-making. Accordingly, either teachers allow the curricular pacing of their instruction to be regimented by what's to be assessed on these interim tests (and few teachers relish such regimentation), or teachers will find their instruction is out of line with what's being tested. Perhaps this is why no evidence regarding the profound payoffs of interim tests has yet been seen. Perhaps, for most teachers, interim assessments just don't work.

Conclusion

Nonetheless, we continue to see ardent advocacy for the installation of balanced-assessment approaches. Much of this advocacy can be traced back to the very folks who sell such tests. If balanced assessment comes to be seen as *necessarily* including interim assessments, then those who sell such assessments can be assured of a serious slice of assessment's fiscal pie. Yet, until suitable support for interim tests arrives, balanced assessment will most definitely remain out of balance.

About the author

W. James Popham is Professor Emeritus at UCLA Graduate School of Education and Information Studies. He spent the bulk of his educational career as a school teacher, and later taught courses in instructional methods for prospective teachers as well as courses in evaluation and measurement for graduate students at UCLA. He has written and co-written 30 books, 200 journal articles, 50 research reports, and 175 papers.

Further reading/resources

Arter JA (2010). *Interim Benchmark Assessments: Are We Getting Our Eggs in the Right Basket?* Paper presented at the annual meeting of the National Council on Measurement in Education, Denver www.assessmentinstitute.com.

Heritage M (2010), *Formative Assessment: Making It Happen in the Classroom*. Corwin Press: Thousand Oaks, CA.

Heritage M (2010), *Formative Assessment and Next-generation Assessment Systems: Are We Losing an Opportunity?* Council of Chief State School Officers: Washington, DC. www.ccsso.org/resources/publications.html?name-search=Heritage+A

Popham WJ (2010), *Everything School Leaders Need to Know About Assessment*, Thousand Oaks, CA: Corwin Press.



Do we need an assessment overhaul?

Jay McTighe and Grant Wiggins suggest a different approach to measuring students' progress

THE EMERGENCE OF the new Common Core Standards presents an opportunity to re-examine the current system of educational assessments in the U.S. For the past 10 years, the No Child Left Behind (NCLB) federal statute has required annual state testing as a means of gauging student achievement. Publishing these test scores

establishes accountability, comparing schools and districts, and resulting in consequences for schools that fail to achieve “annual yearly progress” quotas. Responsible educators understand the need for accountability and the NCLB testing program has revealed achievement deficiencies that demand to be addressed.

Nonetheless, the present assessment system is flawed, and ironically may impede the very efforts needed to attain important educational goals.

The adage, “What gets measured signals what is important,” rings true in education. Students regularly ask their teachers, “Will this be on the test?” If the answer is “No,” they are less likely to pay attention. Large-scale assessments naturally hold even greater sway. Teachers and administrators pay close attention to state and provincial assessments since their results can have high stakes consequences, not only for students but for schools. If something is not assessed, it can receive less emphasis in the classroom. The result is often a *de facto* narrowing of the curriculum, and misguided “test prep” interventions.

Currently, NCLB employs a “snapshot” approach to assessment through annual state testing in targeted subject areas. Given the large-scale nature of these tests, the majority of them understandably employ a selected-response format, allowing for fast, inexpensive, machine scoring. This type of assessment is simply incapable of measuring students’ responses to open-ended problems and issues, discussion and debate, extended writing for real audiences, or showing substantive research and experimental inquiry. Moreover, many subject areas for which standards exist are not tested at all in many states, and nor do these accountability measures typically test the so-called 21st-century skills of creative thinking, teamwork, multi-media communication, and use of information technologies. It can be argued that current standardized assessments fail to assess many of the most valued goals of schooling.

An alternative approach

We recommend an alternative approach to assessment that can minimize unhealthy curriculum narrowing, provide more robust evidence of academic knowledge and 21st-century outcomes, and support meaningful learning through authentic and engaging teaching. Our framework offers a viable approach for achieving three inter-related goals:

- Assessing the most important educational goals in appropriate ways;
- Providing the specific and timely feedback needed to improve learning; and
- Supporting curriculum planning, local assessment, and teaching for meaningful learning.

To achieve these goals, we propose a “multiple measures” approach, with three components for assessing core requirements and other important educational outcomes.

ASSESSMENT

Time for an overhaul

These are:

- Content-specific tests;
- A series of content-specific and interdisciplinary performance tasks; and
- A local assessment component.

Content-specific tests

Content-specific tests, consisting of multiple choice and brief constructed response (BCR) items designed to test core requirements, do have a value. These types of test have been proven to be effective and efficient at sampling a broad array of basic knowledge and skills drawn from the curriculum. These tests should be computer-based in order to take advantage of enhanced item types made possible through technology-enabled assessments, and to provide nearly immediate feedback in the form of detailed item analyses (not just scores).

We further propose that a Matrix sampling approach be considered as a cost-saving means of obtaining accountability information at the school and district levels without subjecting every student to testing every year on every aspect of the Core Standards. However, states or school districts could opt for census testing if individual student scores are desired. Of course, this type of testing is limited, and therefore needs to be accompanied by other types of test.

Content-specific and interdisciplinary performance tasks

Performance tasks call for students to apply their learning to new situations in context. Accordingly, they are better suited to assess more complex concepts and

What we know

- Current standardized assessments do not adequately assess many of the most valued goals of schooling.
- A better “multiple measures” approach would include content-specific tests; a series of content-specific and interdisciplinary performance tasks; and a local assessment component.
- This would help schools to test what they teach, rather than teaching to the test.

would also be a base from which national or regional assessments could be developed.

It is important to note that scoring would not be contracted to commercial test companies, although companies may be enlisted to help with training, moderation, and reporting. Indeed, a central feature of this proposal relates to the high-impact professional development that accrues when teachers work in teams to score students’ work. Accordingly, the costs of scoring the performance tasks need to be conceived and budgeted as a joint expenditure for assessment *and* professional development.

Local assessments

Standardized national assessment systems are incapable of assessing every student on every educational goal. Therefore, the third component of our system legitimizes the role of local assessment, trusting teachers with the responsibility of scoring work in *all* subject areas. The results, framed in terms of a system of standards, would be made public.

standardized assessments;

- Honors the tradition of local control of education by allowing local decision making, rather than having all high-stakes assessments imposed from the outside; and
- Targets student accountability; i.e., the results become part of local grading and reporting.

A cornerstone of this third component is a *Student Standards Folder* – a systematic collection of assessment evidence related to Core Standards and other important educational goals. This would include the results from the performance tasks, content specific tests, and local assessments, as well as rubrics in each subject area to enable more systematic tracking of student achievement (i.e., progress toward meeting standards). The folder would be audited on an annual basis by regional teams of educators and “citizen experts,” with two content areas sampled each year for a state audit. The system would enable educators, parents, and students to track progress over time.

Conclusion

Sadly, the use of classroom time in many schools (at least in the tested grades and subjects) would lead one to conclude that the mission of schools is to improve test taking savvy and raise test scores, rather than to strive for meaningful learning. Of course, it makes sense to familiarize students with test format, but excessive test preparation is *not* the best long-term strategy for developing a well-rounded, educated person *or* improving scores on yearly accountability tests. We contend that our three-part system provides a more comprehensive system for assessment, while avoiding some of the problems of current accountability testing.

About the author

Jay McTighe is an Educational Consultant. With an extensive background in professional development, he is a regular speaker at national and international conferences and workshops. He has co-authored 11 books, including the best-selling *Understanding by Design* series with **Grant Wiggins**, and written more than 30 articles and book chapters. Jay earned his Masters degree from The University of Maryland and has completed post-graduate studies at Johns Hopkins University.

Further reading

Wiggins G and McTighe J, (2007) *Schooling By Design: Mission, Action, Achievement*. Alexandria: ASCD

☞ The adage, ‘What gets measured signals what is important,’ rings true in education ☞

21st-century skills, such as mathematical reasoning, scientific investigation, issues analysis, creative problem solving, oral communications, and technology applications. Performance assessments should be set in real-world contexts, and include both content specific and interdisciplinary performances. Importantly, they should be implemented by teachers *as part of the curriculum* at designated time periods during the school year. Other nations (e.g., the UK) already include assessments scored by teachers as a major element of their national assessments.

This type of assessment would ideally be translated into a national database of performance tasks and companion scoring rubrics, accessible to all teachers so that ideas and resources can be shared. This

The local component of the assessment system allows for a wide variety of possibilities, including common course exams, student projects and exhibitions, and interdisciplinary tasks involving collaboration and technology applications. More specifically, it:

- Can appropriately assess important achievement targets (e.g., oral reading and speaking, applications of technology, teamwork) that may otherwise “fall through the cracks;”
- Is based on local curricula so that teachers, students, and parents will be more likely to “own” the measures and the results;
- Offers greater flexibility and potential for differentiation (e.g., giving students some choice of topics or products) than

Formative assessment: An enabler of learning

Formative assessment can be a powerful day-to-day tool for teachers and students. **Margaret Heritage** explains

FORMATIVE ASSESSMENT is often misconstrued. Routinely, it is conceptualized as a “test” or an “instrument” that is more fine-grained and administered more frequently than other types of assessment. This formulation misses its documented power for improving student learning. When formative assessment is conceived as a practice implemented by teachers, in collaboration with their students, then its promise as an enabler rather than an evaluator of learning can be realized.

The essential purpose of formative assessment as a practice is to move students’ learning forward while their learning is still in the process of developing. This stands in contrast to other forms of assessment, which evaluate learning after a period of teaching. Formative assessment practice operates as a feedback loop in which both teachers and students can play active, distinctive, yet complementary roles in enabling learning by consistently working to build and consolidate student understanding and skills during the course of a lesson.

The teacher’s role

Formative assessment is only effective when teachers are clear about the intended learning goals for a lesson. This means focusing on what students will learn, as opposed to what they will do, which is often where teachers are tempted to start. To achieve maximum transparency for students, teachers share the learning goal, or actively create it with students, at the beginning of the lesson. In addition, teachers communicate the indicators of progress toward the learning goal or determine them in collaboration with the students. These indicators serve as signposts for both teachers and students about progress during the lesson.

With clarity about the goal and indicators, teachers can then decide how they will gather evidence of emergent learning. There is no single way to collect formative evidence because formative assessment is not a specific kind of test. For example, teachers can gather evidence through interactions with students, observations of their tasks and activities, or analysis of their work products. However, there are two important points about evidence collection. First, whatever method teachers use to elicit evidence of learning, it should yield information that is actionable by them and their students. Second, evidence

Power and responsibility in the classroom is not just the teacher’s prerogative, but is distributed

collection is a systematic process and needs to be planned so that teachers have a constant stream of information tied to indicators of progress. At the same time, of course, teachers will also be collecting evidence “on-the-fly” – those unplanned, spontaneous moments when students do or say something that give an indication of where they are in relation to the lesson goal.

Feedback

Feedback is a crucial component of formative assessment, and has two aspects. First, feedback obtained from planned or spontaneous evidence is an essential resource for teachers to shape new learning through adjustments in their instruction. If teachers use evidence effectively to inform their instruction, it will render previous



assessment information out of date: student learning will have progressed and will need to be assessed again. Instruction can again be adjusted to make sure that learning is on track. For this reason, a constant stream of evidence from formative assessment is necessary during lessons.

Second, feedback that the teacher provides to students is also an essential resource so the students can take active steps to advance their own learning. In reality, the feedback to students can be understood as instructional action. As the extensive literature on feedback suggests, teacher feedback is most beneficial when it assists students to understand their current learning status and provides hints, suggestions, or cues for them to act on. It is this, rather than offering general praise or total solutions, that enables students to assume a degree of responsibility for their learning.

The teacher’s role also involves helping students develop the skills to make metacognitive judgments about their learning in relation to the goal being aimed for, and to establish a repertoire of strategies to regulate their own learning.

The students’ role

The students’ role in formative assessment begins when they have a clear conception of the learning target. Just as the teacher



What we know

- Formative assessment is not a kind of test.
- Formative assessment practice, when implemented effectively, can have powerful effects on learning.
- Formative assessment involves teachers making adjustments to their instruction based on evidence collected, and providing students with feedback that helps them advance their learning.
- Students participate in the practice of formative assessment through self- and peer-assessment.

is collecting evidence in relation to the goal, so too are the students through self-assessment, a separate, but complementary feedback process. In self-assessment students engage in metacognitive activity, a hallmark of effective learning. Metacognitive activity involves students in thinking about their own learning while they are learning. In this process, they are generating internal feedback that tells them when they need to make adjustments to their learning strategies. These adjustments might include, for example, drawing a diagram to help in the understanding of a mathematical problem,

or determining that more research is needed to be able to analyze historical events, or re-reading a text to clarify the meaning.

The students' role ideally also includes peer-assessment. In peer-assessment, students give feedback to their classmates that is intended to be constructive and help them make progress toward the lesson goal. Peers assess each other's learning against the same indicators that they use to check on their own learning when they are engaged in self-assessment. Peer feedback has a number of advantages both for those students providing the feedback and for those receiving it. It involves thinking about learning and can deepen students' understanding of their own learning because they have to internalize the learning goal and progress indicators in the context of someone else's work.

The final point about the students' role in formative assessment is that they actually use the feedback. It is important that students have to both reflect on their learning and use the feedback to advance learning. One teacher summed up the changes she made to ensure that this time was preserved in her lessons:

"I used to do more but now I do less. Now I work hard to save time for student reflection rather than filling every minute [of the lesson] with activity."

Overall, the feedback loop is fueled by three convergent sources of feedback: from teachers, peers, and the students themselves. However, the successful provision and use of this feedback is dependent on the nature of the classroom climate in which the learning is taking place.

Classroom climate

An essential aspect of formative assessment is classroom climate. Three particular elements are key. First, power and responsibility in the classroom is not just the teacher's prerogative, but is distributed so that teachers and students work together to share responsibility for learning. Second, the classroom has to be a safe place. Students must be able to ask for help, regard errors as sources of new learning, and admit difficulties or problems without fearing that these actions will diminish them in the eyes of their teachers or their peers. Instead, they need to know that such behaviors are desirable and are characteristic of effective learners. Finally, it means that the relationships in the classroom must be supportive and collaborative, characterized by mutual trust among teachers and students.

Conclusion

The important thing about formative assessment is that it is not a test, nor an instrument, but rather an approach to teaching and learning that uses feedback as its centerpiece in a supportive classroom context. Formative assessment is a practice that empowers teachers and students to give their best to enable learning. In the end, the success of formative assessment as an enabler of learning depends on the knowledge and skills of teachers to implement this approach in collaboration with their students, not on test developers.

About the author

Margaret Heritage is Assistant Director for Professional Development at the National Center for Research on Evaluation, Standards and Student Testing (CRESST) at the University of California, Los Angeles. She has written and co-written numerous articles and books on formative assessment.

Further reading

Black PJ and Wiliam D (1998), Assessment and Classroom Learning. *Assessment in Education: Principles, Policy and Practice*, 5, 7–73.

Hattie J and Timperely H (2007), The Power of Feedback. *Review of Educational Research*, 77, 81–112.

Upgrading high-stakes assessments



Albert Oosterhof describes current research into the effectiveness of a new assessment strategy

HIGH-STAKES ASSESSMENTS have existed for many years; for instance, the ordinance that created the Regents examination system in New York State was passed in 1864. The use of high-stakes assessments has become widespread in recent years, and in the U.S. their prevalence increased with passage of the No Child Left Behind Act of 2001. As the term “high stakes” suggests, students’ performance on these assessments can result in significant actions directed at students, teachers, and/or schools.

Information derived from high-stakes assessments is used summatively, not formatively. Because of the large number of students involved, the assessments must be highly efficient with regard to administration and scoring. This constrains the formats that can be used and limits testing to a subset of competencies typically associated with the standards being assessed. Teachers are motivated to emphasize this subset to help their students perform well on the tests. This narrowing of the curriculum is likely to become more significant as state governments consider legislation that links teachers’ salaries to test scores.

Nevertheless, by using carefully constructed performance assessments, it is possible to assess important skills that are excluded from high-stakes tests. The problem is that the cost per student to administer and score these complex assessments is high.

A new assessment strategy

Through a three-year grant from the U.S. Department of Education, the Center for Advancement of Learning and Assessment at Florida State University will examine the feasibility of a three-pronged strategy that may:

- Help expand the range of skills evaluated by statewide assessments; and
- Add a formative aspect to these assessments.

The three components are as follows:

Performance assessments: The first aspect involves developing a series of performance assessments that measure selected state-level benchmarks and are administered to carefully selected samples of students. This part of the strategy estimates student proficiency at the group level and, like the National Assessment of Educational Progress (NAEP), is not designed to determine individual student proficiency. The content and number

of performance assessments are controlled to provide appropriate generalizability, with emphasis placed on competencies that cannot be effectively measured using conventional assessments. Using sampling would be less expensive than testing every student.

Measuring complex skills: The second component focuses on the proficiency in complex skills of each individual student. It entails developing performance assessment “specifications” that define comparable measures to be developed by teachers, linking teachers’ assessments to those administered statewide to samples of students. Students’ performance levels on the teachers’ summative assessments are then compared to performance on the assessments administered statewide to samples of students. Assessments administered to these samples substantiate the reasonableness of student outcomes observed through the assessments of teachers, and vice versa. To be effective, these expanded assessment procedures must facilitate learning rather than increase the burdens placed on teachers and school administrators.

Using performance assessments

formatively: The third component involves using the performance assessments not just summatively, but also formatively as an integral part of instruction. Results on performance-based measures would provide the basis for formative feedback to students throughout the school year. Feedback also would link performance on the task to the broader set of tasks implicit in the performance assessment specification. Teachers’ performance assessments eventually have a summative role, establishing what individual students have learned. However, at the classroom and school levels, the major focus is on their formative role. We will employ what research has identified as “best practices” related to the use of formative feedback to students.

How will we measure effectiveness?

Our research will simulate the above components in a controlled setting. We will produce and administer the external assessments that ultimately would be developed by a state assessment office. These assessments will be administered to students enrolled in participating classrooms, not statewide to samples of

students. We know a NAEP-type approach can estimate achievement of groups of students from samples. We do not know whether separate external- and teacher-developed performance assessments, based on common specifications, can validate each other and help substantiate teachers' summative assessments of individual students.

Our research occurs in the context of science instruction taught at the middle school level. The intent, however, is to establish procedures that are useful at other grade levels and in other subject areas. To maximize the benefits of the research, we believe teachers and school administrators must assume ownership of the ideas and procedures that evolve. Therefore, a partnering relationship between teachers, curriculum specialists, measurement experts, and training specialists will be emphasized.

Measuring cognitive processes

When using authentic performance assessments, it is tempting to assume that we are directly measuring the intended outcomes of learning. Any assessment, however, involves measuring something that cannot be seen, that is, a student's cognitive processes. The performance of a student is only an *indicator* of what the student knows and is thinking.

Knowledge exists at different levels of complexity. We have used categories of knowledge based on those often used in cognitive psychology:

- **Declarative knowledge:** being able to explain things, such as how distillation works.
- **Procedural knowledge:** being able to invoke learned techniques, often in new applications, such as using distillation to separate compounds.
- **Problem solving:** when one has a goal but has not yet identified a means for reaching it. Problem solving uses strategies that rely on declarative and procedural knowledge, such as recognizing if distillation can help identify the presence of a particular chemical in a solution.

What we know

- If assessments de-emphasize particular types of competencies, what is learned also will be de-emphasized.
- For practical reasons, high-stakes tests de-emphasize complex skills. It may be practical to assess these skills if samples of students are used.
- Because tests involve only samples of tasks, students' performance may not generalize to what would have been observed had different, equally appropriate tasks been used.

We are particularly interested in complex skills that cannot be assessed with conventional tests, and have found it easier to address complexity levels after identifying the above categories of knowledge. For instance, a performance assessment involving procedural knowledge might ask a pupil to accurately determine true north, east, south, and west using sun shadows, without reference to other objects that indicate direction. A conventional test would involve a less complex task, as illustrated with this multiple-choice item:

A stake has been placed straight up, on flat ground, in sunlight. When does the stake's shadow fall in a true north and south direction?
A. When the shadow is at its shortest length (correct).

B. When the end of the shadow is moving the fastest (incorrect).

With declarative knowledge, a performance assessment might ask a pupil to explain in writing (1) what you did to find true north, east, south, and west and (2) why your technique works. A less complex declarative task would be presented with this multiple-choice item:

At any location in the United States, when does "local noon" occur?

A. When the sun is exactly south of that location (correct).

B. When the local time zone switches from morning to afternoon (incorrect).

Important issues

Our research into upgrading high-stakes assessments is just beginning. Some important issues exist that we and other researchers working in this area must address.

Any written test or performance assessment involves only a sample of tasks that might have been used for the assessment. An important question is whether the same conclusions regarding student achievement would have been reached had different, equally appropriate tasks been used. If students' performance does not "generalize," the assessment is of limited value because conclusions based on performance depend on what task was sampled. Generalizability of high-stakes assessments must be high and involve facets other than tasks (e.g., samples of raters used to score performance).

The validity of an assessment is a critical issue. Assessments always involve indicators – not direct observation – of knowledge. Validation means establishing a link between the knowledge we seek to assess and the tasks we ask students to complete. Similarly, interpretation of performance requires establishing this link in the opposite direction, between what students were observed doing

and the knowledge being assessed. Evidence-centered design will provide a framework for accomplishing this task.

Scalability and practicality are important issues that must be resolved if we are to successfully implement this alternative approach to high-stakes assessment programs.

Project details

The work reported in this paper is supported through a recently awarded grant from Education Research Programs at the Institute of Education Sciences (IES), award number R305A110121, administered by the U.S. Department of Education. Faranak Rohani, director of the Center for Advancement of Learning and Assessment, is the principal investigator for this research. Related information is available at <http://cala.fsu.edu/ies/>. Findings and opinions do not reflect the positions or policies of the IES or the U.S. Department of Education.

About the author

Albert Oosterhof is Professor Emeritus in Educational Psychology and Learning Systems at Florida State University and a research associate at the university's Center for Advancement of Learning and Assessment. The focus of his work is on student assessment. (albert.oosterhof@fsu.edu)

Further reading

Chi MTH, and Ohlsson S (2005), Complex Declarative Learning. In KJ Holyoak, and RG Morrison (Eds), *Cambridge Handbook of Thinking and Reasoning* (pp. 371–399). New York: Cambridge University Press.

Mislevy RJ, Almond RG, and Lukas JF (2004), *A Brief Introduction to Evidence-Centered Design*, (CSE Report 632). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing. Also available at www.cse.ucla.edu/products/reports/r632.pdf

Oosterhof A et al (2008), *The Capabilities-Complexity Model*. Symposium paper presented at the National Conference on Student Assessment, Orlando, FL. Also available at www.cala.fsu.edu/files/ccm.pdf

Solano-Flores G et al (1997), *On the Development and Scoring of Classification and Observation Science Performance Assessments*, (CSE Technical Report 458). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing. Also available at www.cse.ucla.edu/products/Reports/TECH458.pdf

England v Wales: Education performance and accountability

There is increasing disparity between the achievement of students in England and Wales. Sandra McNally explores the possible causes

IN MANY RESPECTS, the education systems in England and Wales are very similar. They had a shared history within the UK up until Welsh devolution in 1999, and although differences have increased since then, the main examination for 16-year-olds in both countries remains the General Certificate of Secondary Education (GCSE), which awards academic qualifications to students in specific subjects. GCSEs are graded from A* (pronounced as “A star”) to G. In the UK, examinations at 16 are a key educational milestone.

In this article, we consider a few of the differences between education in England and Wales – focusing on accountability in particular. We discuss possible reasons for the difference in educational performance and the potential contribution of differing policies on the publication of “league tables” of examination results.

Comparing educational performance in England and Wales

Looking at GCSE attainment in England and Wales, it is interesting to observe how the two countries used to perform exactly the same in terms of national statistics (at least between 1998 and 2001) with

differences starting to emerge from 2002 onwards – escalating after 2004. The figure below shows the headline statistic in the two countries.

England and Wales also used to have similar assessments for students at the end of primary (elementary) school, and the results are comparable up to about the year 2000. In this case, the two

countries also performed similarly in the mid-1990s (particularly for English) with a divergence emerging in 1998 (for English) and 1999 (for math) – with England taking over.

Are ‘league tables’ responsible?

League tables are produced annually by the UK Government’s Department for Education and currently list statistics for GCSE exam performance by every school in England, as well as several other key indicators. Previously, the tables also included

information for Wales, but this was scrapped by the Welsh Assembly in 2001. This coincides with the date when England and Wales began to diverge in terms of GCSE grades.

League tables continue as a prominent feature of education in England. The rationale is that they provide information to parents on the relative performance of different schools, and that parents can use this information to help them decide where to send their children. Although league tables report more than just raw exam grades, qualitative evidence suggests that school principals focus almost exclusively on the above-reported indicator

It is plausible that at least some of the relative deterioration in performance in Wales compared to England came about because the relative incentive to engage in such practices was diminished in Wales

(5 or more “good grades”). Also, quantitative research suggests that raw grades are important in terms of the premium paid on house prices for living near a “good school.”

Could the abolition of league tables in Wales somehow be responsible for this diverging performance with England? And what would this mean? First take the case where we really believe the GCSE indicator is an accurate measure of how much knowledge and skills students acquire at school. Then the divergence with England is most definitely a bad thing and the abolition of league tables in Wales could be responsible to the extent that school principals stop making such an effort because outcomes are less visible.

Funding and policy differences

On the other hand, there are other things that might explain the divergence. For example, it is well known that average funding per student is much lower in Wales. Currently, funding per student is about 10% lower in Wales than in England.

Furthermore, many policy initiatives introduced in England have not been introduced in Wales since devolution. It is not clear *a priori* whether this is a good or a bad thing. However, some policies have been properly evaluated. One of the big differences between England and Wales is that the national Literacy and Numeracy Strategies that were introduced in England in the late 1990s were not introduced in Wales. The Strategies involved England adopting a very

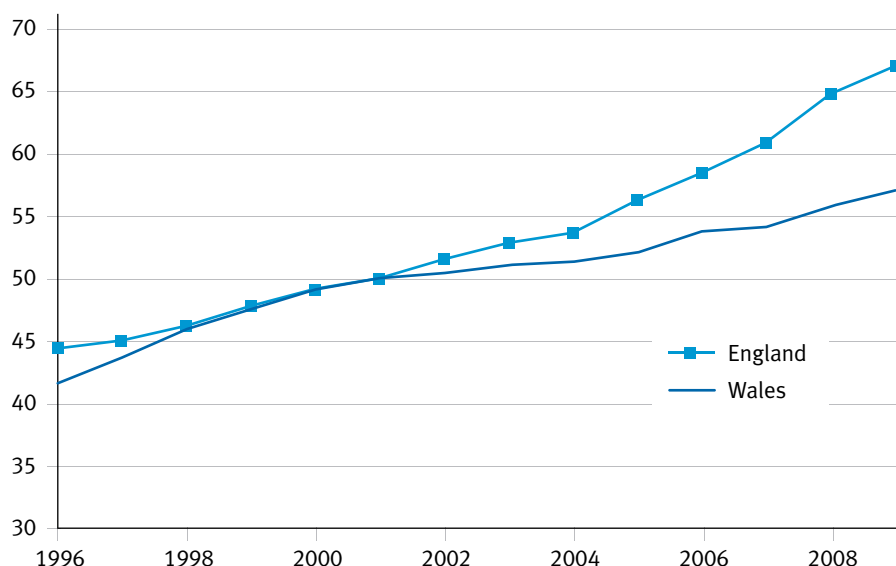


Figure 1. Percentage of students achieving five or more GCSEs with a grade of A* to C

centralized approach involving literacy and numeracy hours that had to be implemented in primary schools. These were quite prescriptive in terms of the pedagogy. Policy makers in Wales rejected the “top-down” approach and instead encouraged local authorities and schools to develop their own, locally based initiatives. Research evidence shows that diverging attainment at age 11 between England and Wales is associated with the introduction of the Strategies in England, even when taking account of other school-level characteristics. The first cohorts exposed to the Literacy and Numeracy Strategies in England would have taken their GCSE examinations in 2003 and 2004 respectively. It is around this time that the divergence between England and Wales really starts to escalate.

There is only so much one can infer from looking at trends in raw results between countries. There are at least three competing explanations for the apparent deterioration in Wales – the abolition of league tables, lower per student funding, and the decision not to implement National Literacy and Numeracy Strategies in Wales.

Are high stakes tests the best measure of student knowledge?

Another question is whether the divergence between England and Wales after 2001 in GCSE attainment really reflects differences in the extent to which students have appropriated knowledge and skills. It is often argued that “high-stakes” tests encourage various teacher practices designed to boost results, but not necessarily knowledge and skills. These practices include:

- Coaching students to respond to test questions rather than teaching them mastery of a subject;
- Narrowing their focus to cover only those topics that will be assessed;

What we know

- Educational performance has lowered in Wales compared to England in recent years.
- The timing coincides with the scrapping of publication of exam results in Wales.
- Other potential explanations are lower funding per student in Wales and the success of the Literacy and Numeracy Strategies in England.
- The abolition of “league tables” of examination results in Wales gave teachers less incentive to focus narrowly on assessment.
- International low-stakes tests also suggest a relative decline in Wales.



- Focusing on students near the threshold of the headline indicator at the expense of other students;
- Encouraging students to take “easy” subjects; and sometimes
- Outright cheating (e.g., misuse of continuous assessment).

There have been many international studies that find evidence of these practices in cases where high-stakes testing is introduced. One of the ways this is shown is by comparing results in high-stakes tests versus low-stakes tests for the same students. It has often been found that the improvement in high-stakes tests (like GCSE) is not reflected in low-stakes test. Such considerations have led to recommendations that education authorities separate assessments that aim to measure educational progress from assessments that aim to incentivize schools and teachers. (This is a recommendation of Professor Derek Neal of the University of Chicago, who has recently produced a review of evidence on the design of performance pay in education.)

It is plausible that at least some of the relative deterioration in performance in Wales compared to England came about because the relative incentive to engage in such practices was diminished in Wales.

However, there is a low-stakes test to which we might refer. Both England and Wales participated in the OECD international test for 15-year-olds (PISA) in 2006 and 2009. This could be thought of as low stakes because school performance in PISA is not made publicly available. In both years, average performance in Wales was lower than that in England. Furthermore, comparing 2006 and 2009, performance decreased in reading, mathematics, and science in Wales whereas there was little change in England. This suggests the increasing gap between

England and Wales is not just an artifact of the British examination system or the incentives produced by differing policies about league tables. One might debate the extent of the difference, but the existence of the phenomenon seems to be well-founded. Possible reasons merit further investigation.

About the author

Sandra McNally is a Senior Research Fellow at the Centre for Economic Performance at the London School of Economics. She is director of the Education and Skills program.

Further reading

BBC News: *PISA Tests Show Students in Wales Falling Further Behind*, 7 December 2010.

www.bbc.co.uk/news/uk-wales-11930257

Bradshaw J, Ager R, Burge B and Wheeler R (2010), *PISA 2009 Achievement of 15-year-olds in England*. National Foundation for Educational Research. Slough. England.

Gibbons S, Machin S and Silva O (2009), *Valuing School Quality Using Boundary Discontinuity Regressions*. SERC Discussion Paper 18. London School of Economics.

Neal D (2011), *The Design of Performance Pay in Education*. NBER Working Papers. Cambridge Massachusetts.

Machin S and McNally S (2009), *The Three Rs: What Scope is there for Literacy and Numeracy Policies to Raise Pupil Achievement?* Mimeo. Centre for Economic Performance, London School of Economics.

Wilson D, Croxson B and Atkinson A (2006), *What Gets Measured Gets Done*. *Policy Studies*. 27(2). 153–171(19).

Johns Hopkins researchers examine effectiveness of education technology

ACCORDING TO A COMPREHENSIVE research review from Johns Hopkins University's Center for Research and Reform in Education (CRRE), education technology generally produces a positive, though small, effect on K–12 reading achievement. However, the effects may vary by education technology type.

In order to identify relevant studies for their review, CRRE researchers conducted a literature search to locate all studies of technology-assisted reading interventions that were completed between 1970 and 2010. To be included in the review, studies had to meet a set of rigorous inclusion standards. For example, studies had to use randomized or matched control groups, have a duration of at least 12 weeks, and use valid achievement measures that weren't inherent to the program being tested. A total of 85 studies based on 60,712 K–12 participants qualified and were included in the review.

The researchers grouped the qualifying studies by intervention type into four main categories: computer-managed learning, innovative technology applications, comprehensive models, and supplemental technology. The majority of the studies (67%) fell into the supplemental technology category.

CRRE suggests that research dollars continue to be invested in the evaluation of education technology

After a thorough analysis of the studies and their respective findings, the researchers concluded that supplemental technology programs (i.e., programs that provide additional instruction at students' assessed levels of need) have the least evidence of effectiveness out of the four categories reviewed and are not

producing an educationally meaningful impact on K–12 reading achievement. In contrast, computer-managed learning (e.g., *Accelerated Reader*), innovative technology applications (e.g., *Lightspan* and *Reading Reels*), and comprehensive models (which combine computer and non-computer classroom instruction with extensive professional development support) did show promising effects. The researchers caution, however, that too few rigorous studies for these promising approaches are available at this point for firm conclusions.

Based on these findings, CRRE suggests that research dollars continue to be invested in the evaluation of education technology. Technology's role in the K–12 classroom will undoubtedly continue to expand, so it is important for educators to know what works.

Johns Hopkins University Best Evidence Encyclopedia, January 2011
www.bestevidence.org/reading/tech/tech.html

Study shows that pre-k attendance can lead to strong academic gains

A STUDY FROM the Peabody Research Institute at Vanderbilt University finds that the Tennessee Voluntary Pre-K Program – a state-funded initiative which seeks to provide four-year-old children with the learning experiences they need in order to succeed in kindergarten – is dramatically improving children's early literacy, language, and math skills.

Included in the study were 303 children from 23 schools in Tennessee. These schools did not have enough spaces for all of the children who applied to their pre-k classes, so they admitted students in the order prescribed by a list that randomly scrambled the sequence of names. This allowed researchers to conduct a randomized controlled trial of the program, in which a randomly-assigned treatment group (the students who were admitted) was compared with a randomly-assigned control group (the students who applied but were not admitted).

According to the study, 230 children were accepted into the



schools' pre-k classes for the 2009–2010 school year. These children's literacy, language, and math skills were assessed at the beginning and end of the year and compared with the skills of the 73 children in the sample who were not admitted to the pre-k classes. Standardized instruments from the Woodcock Johnson III Battery were used to measure the children's skills.

Results of the study showed that the treatment group's skills improved an average of 82% more than the control group. The strongest gains were in literacy and language, but significant gains were also shown for math. However, these results reflect fairly immediate effects after only one year. Importantly, children in the study will be followed for a total of five years (into the third grade) to determine lasting impacts.

**Vanderbilt University
Peabody Research Institute
February 2011**
news.vanderbilt.edu/2011/02/tennessee-pre-k-students-see-82-percent-gain-over-peers



Issue brief identifies promising practices for promoting high school graduation

IN A LABOR MARKET where education is typically linked with greater job opportunity, high school dropouts can find themselves at a distinct disadvantage. In addition, research has shown that educational attainment is positively associated with outcomes such as better health. To help educators keep all students on track to graduate, the RAND Corporation's Promising Practices Network (PPN) developed an issue brief that describes multiple, research-based approaches for promoting high school graduation. To be included in the brief, approaches had to meet PPN's "promising" level of evidence at a minimum (see evidence criteria at www.promisingpractices.net/criteria.asp).

The brief first discusses interventions that have been shown to promote school retention among at-risk students by focusing on improving academic performance. According to the brief, these approaches include intensive instruction in particular subjects, such as math or reading, or personalized instruction, which may include extra instructional time or help students address specific academic performance challenges, such as poor test-taking or study skills.

Approaches that focus on non-academic skills (e.g., social interaction and decision making) are also discussed. Specifically, the brief identifies the following as approaches with evidence of effectiveness: mentoring

programs, pairing students with counselors who regularly meet with them, and placing at-risk students in smaller classrooms where the learning environment is more personalized and teachers become more familiar with students' strengths and needs.

High-quality, intensive early childhood education is also identified as a promising practice for promoting high school graduation. According to the brief, early education may promote later educational outcomes by improving both students' academic and non-academic skills.

Promising Practices Network
February 2011

www.promisingpractices.net/briefs/briefs_highschoolgrad.asp

Report analyzes common practices among improving school districts around the world

HOW DOES A SCHOOL SYSTEM with poor performance become good? And how does one with good performance become excellent? Those are two questions examined by McKinsey & Company – an international consulting firm – in a report they completed as a follow-up to their 2007 publication, *How the World's Best-Performing School Systems Come Out on Top*.

For their most recent report, researchers analyzed 20 school districts from around the world that were identified as having achieved significant, sustained, and widespread gains in student outcomes on international and national assessments from 1980 onwards. The researchers looked at the different reform efforts carried out by each system in an effort to identify common reform elements that have led to increased student achievement. School systems from five continents were included in the sample.

To gain a better understanding of all the reform interventions that had been implemented in each system, the researchers conducted interviews with more than 200 system leaders and their staff. Based on the information shared during these interviews, the researchers created a database of nearly 575 individual interventions that they identified across the 20 systems. The researchers grouped these interventions into 10 broad areas of impact (e.g., professional development and accountability) and 60 subareas. For example, the broad area of



"accountability" included the subareas of performance assessment, school inspections, and self-evaluation.

Several important findings emerged from the research. For instance, it showed that all improving school systems in the sample implemented similar sets of interventions to move from one particular performance level (e.g., poor performance) to the next (e.g., fair performance), irrespective of culture, geography, politics, or history. Specifically, the report says that systems moving from fair performance to good focused on establishing the foundations of data gathering, organization, finances, and pedagogy, while systems on the path from good performance to great focused on shaping the teaching profession such that its requirements, practices, and career paths are as clearly defined as those in medicine

and law. In addition, the report notes that six interventions were common to all performance stages across the entire improvement journey:

- Building the instructional skills of teachers and the management skills of principals;
- Assessing students;
- Improving data systems;
- Facilitating improvement through the introduction of policy documents and education laws;
- Revising standards and curriculum; and
- Ensuring an appropriate reward and remuneration structure for teachers and principals.

Further findings and in-depth analysis are included in the full report.

McKinsey & Company, November 2010

www.mckinsey.com/clientervice/Social_Sector/our_practices/Education/Knowledge_Highlights.aspx

The Latest Research

Report: *Effective Reading Programs for Title I Schools*. White paper from Johns Hopkins University. (February 2011)

What? What works in teaching reading? To find out, this white paper examines research on the most effective reading programs for students in elementary school. The purpose of the paper is to identify proven strategies for Title I schools.

Three systematic reviews conducted by Johns Hopkins University's Center for Research and Reform in Education are discussed: (1) *Effective Beginning Reading Programs: A Best-Evidence Synthesis*, (2) *Beyond the Basics: Effective Reading Programs for the Upper Elementary Grades*, and (3) *Effective Programs for Struggling Readers: A Best-Evidence Synthesis* (all available at www.bestevidence.org). For each review, study inclusion criteria included the use of randomized or matched control groups, study duration of at least 12 weeks, and valid achievement

measures independent of the experimental treatments. A total of 63 beginning reading (grades K or 1), 80 upper elementary reading (grades 2-5), and 96 struggling readers studies met these criteria and were included in the reviews.

For beginning reading and upper elementary reading, four types of approaches to improving the reading success of non-struggling students were examined: reading curricula, instructional technology, instructional process programs, and combinations of curricula and instructional process. According to the paper, key findings from the reviews showed that instructional process programs designed to change daily teaching practices have substantially greater research support than programs that focus on curriculum or technology alone.

In the struggling readers review, four types of approaches to improving the reading success of struggling students in grades K-5 were examined: one-to-one

tutoring, small group tutorials, classroom instructional process approaches, and computer-assisted instruction. The paper identifies several important findings that emerged from the research:

- One-to-one tutoring is very effective in improving reading performance;
- Tutoring models that focus on phonics obtain much better outcomes than others; and
- Instructional process programs, especially cooperative learning, can have very positive effects for struggling readers.

Taken together, these findings support a strong focus in Title I schools on improving classroom instruction and then providing one-to-one, phonetic tutoring to students who continue to experience difficulties.

Author: Slavin R.

Where? The white paper can be found at www.bestevidence.org/word/white_paper_Feb_2011.pdf

Report: *Using Student Performance to Evaluate Teachers*. Research Brief from the RAND Corporation. (February 2011)

What? This research brief presents five educational policy recommendations regarding using student performance to evaluate teachers. The recommendations are drawn from a RAND Corporation report that reviewed literature on the topic and profiled three districts and two states that have begun or are planning to incorporate measures of student performance into their teacher evaluations. The locations that were profiled were Denver, Colorado; Hillsborough County, Florida; Washington, D.C.; and the states of Tennessee and Delaware. The purpose of the report was to examine how these five educational systems have already begun or are planning to address challenges associated with using student performance to evaluate teachers. For example, one challenge identified in the brief is how to generate valid estimates of teachers' contributions to student learning.

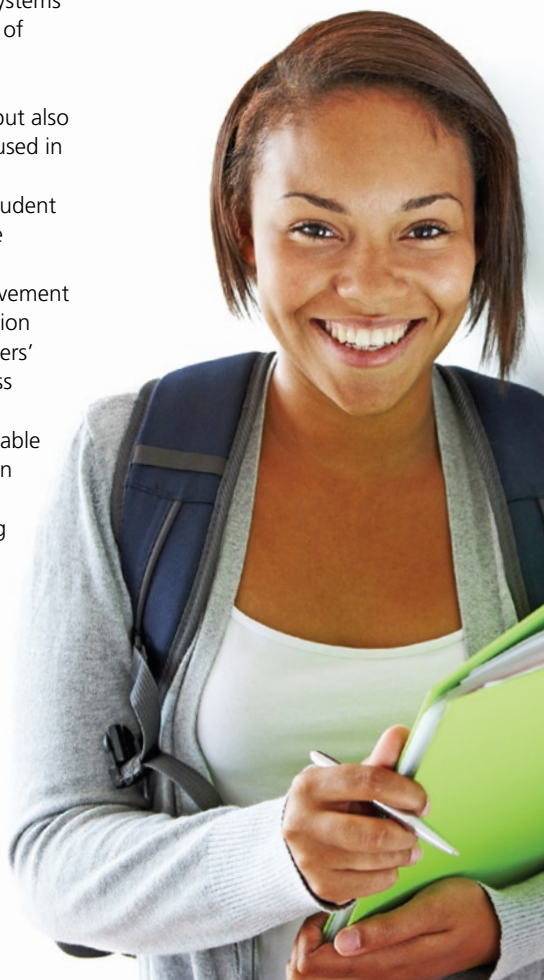
Based on the literature review and insights gleaned from the five profiles, RAND researchers developed several recommendations for policy makers and practitioners who are working to develop new teacher evaluation systems. These recommendations – which are expanded on in the brief – are as follows:

- Create comprehensive evaluation systems that incorporate multiple measures of teacher effectiveness.
- Attend not only to the technical properties of student assessments but also to how the assessments are being used in high-stakes contexts.
- Promote consistency in whatever student performance measures teachers are allowed to choose.
- Use multiple years of student achievement data in annual value-added estimation and, where possible, average teachers' annual value-added estimates across multiple years.
- Find ways to hold teachers accountable for students who are not included in their value-added estimates.

The brief concludes that incorporating student performance measures into teacher evaluation systems is a complex undertaking and, as they strive to enhance teacher evaluation systems, policymakers may benefit from examining what other systems are doing and learning from their struggles and successes.

Author: Newberry S.

Where? The research brief can be found at http://www.rand.org/pubs/research_briefs/RB9569.html



The Latest Research

Report: *Developing Effective Fractions Instruction for Kindergarten through 8th Grade: A Practice Guide* (NCEE #2010-4039). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. (September 2010)

What? This practice guide presents five recommendations intended to help educators improve students' understanding of, and problem-solving success with, fractions. Each recommendation was developed by a panel of researchers and practitioners who have expertise in different aspects of fractions instruction. The panel members worked collaboratively to develop the recommendations based on the best available research evidence and their

combined experience and expertise regarding mathematics teaching and learning. Their five recommendations are as follows:

- Build on students' informal understanding of sharing and proportionality to develop initial fraction concepts.
- Help students recognize that fractions are numbers and that they expand the number system beyond whole numbers. Use number lines as a central representational tool in teaching this and other fraction concepts from the early grades onward.
- Help students understand why procedures for computations with fractions make sense.
- Develop students' conceptual understanding of strategies for solving ratio, rate, and proportion problems

before exposing them to cross-multiplication as a procedure to use to solve such problems.

- Professional development programs should place a high priority on improving teachers' understanding of fractions and of how to teach them.
- Provided in the guide are detailed examples of how to carry out each recommendation. The intended audience for the guide includes elementary and middle school teachers, mathematics supervisors, specialists, coaches, and principals.

Authors: Siegler et al.

Where? The practice guide can be found at whatworks.ed.gov/publications/practiceguides

Report: *What Works for Promoting and Enhancing Positive Social Skills: Lessons from Experimental Evaluations of Programs and Interventions*. Factsheet from Child Trends. (March 2011)

What? This factsheet reviews findings from 38 experimental evaluations of programs that seek to promote positive social skills – such as getting along with others and trying to resolve conflicts – among children and adolescents. The goal of the review was to identify strategies and practices that work, or do not work, in promoting positive social behavior.

The programs' effects on the following five social outcome areas are examined: conflict-resolution skills (e.g., dealing with teasing), interpersonal skills (e.g., sharing), relationship-building skills (e.g., interacting

with others), problem-solving skills (e.g., asking for help), and self-regulation skills (e.g., impulse control). Based on the findings from the experimental evaluations, the programs were grouped into three categories: not proven to work, mixed findings, and found to work. Programs in the "found to work" category had positive and statistically significant impacts on at least one targeted social outcome.

Of the 38 programs included in the review, 27 programs were found to work, eight had mixed findings, and three had no statistically significant impacts. These results suggest that a variety of strategies and practices can be implemented in schools to promote and enhance positive social skills. In particular, the factsheet notes that interventions that incorporated peer-teaching, group discussions, or role modeling, as well as

teacher-led instruction, were effective. Also, programs that combined coaching and/or modeling along with lectures were found to have positive impacts.

As for what did not work, the factsheet does not provide any conclusions. The three programs that had no significant impacts on any social skills outcome varied on targeted population, duration, approach, and format, so the factsheet suggests that further research is needed to better understand which program components lead to unsuccessful results.

Authors: Bandy T & Moore K.

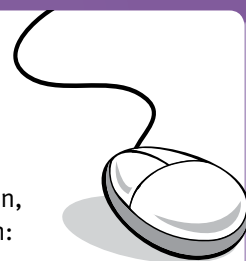
Where? The factsheet can be found at www.childtrends.org/Files/Child_Trends_2011_03_02_RB_WWSocialSkills.pdf



www.bestevidence.org

What works in education? Visit the Best Evidence Encyclopedia (BEE) website to see the latest research. Developed by Johns Hopkins University's Center for Research and Reform in Education, the BEE provides educators, researchers, and policy makers with:

- Free, *Consumer Reports*®-style reviews of reading, math, ELL, and other programs for pre-k to 12 students
- Tools to support school improvement teams
- Resources for implementing research-proven programs



Great educators get **Better**

***Better: Evidence-based Education* tells you what works in teaching.**

Experts in the field write about the state of the evidence on a particular theme - the next issue is on struggling reading. The focus is on practical, evidence-based advice for decision makers in education.

Published three times a year, *Better* has a unique international perspective, with editions of the magazine for both the U.S. and the UK. Articles are written by leading educational researchers and policy makers. Past issues have focused on reading, math, social-emotional learning, science, and effective uses of technology in education.

For more information on *Better*, including how to receive back issues and future copies, visit www.bestevidence.org/better.

